Unresolved Problems in Vadose Zone Hydrology and Contaminant Transport

William A. Jury and Zhi Wang

University of California, Riverside, CA 92521

Despite a great deal of research, many unsolved problems remain in vadose zone hydrology. Because of the difficulties in sampling spatially variable soil and in monitoring at great depths, the effect of heterogeneities on flow and transport is poorly understood. Observations from the few comprehensive field studies that have been conducted suggest that preferential flow, or rapid flow through a small part of the soil volume, is present under a variety of conditions and can be a significant or even dominant part of the total flow regime. Reasons for the occurrence of preferential flow are soil dependent, but include flow through structural voids, channeling caused by discrete obstacles within the soil matrix, and fluid instability (arising from a host of causes). Once believed to occur exclusively in structured soils, preferential flow is now recognized as prevalent under a wide range of conditions in permeable, structureless soils as well as those containing cracks and crevices. Monitoring of preferential flow has thus far only been possible using dyes, densely replicated soil coring, or analysis of tile-drain effluent. These tools either fail to characterize the speed of preferential flow, as is the case with dyes or coring, or provide no spatial resolution of the preferential flow event, as with tile-drain monitoring. To make any real progress in characterizing preferential flow, we need rapid-response tensiometers and solution samplers that offer minimum disturbance of the soil, and we need some means of monitoring a substantial portion of the soil volume at a high degree of spatial resolution.

1. INTRODUCTION

The vadose zone is an important region to a number of different research and applied disciplines. It is the crop-management zone where plants receive nutrients and water through their roots. It is both a waste repository site and an accidental recipient of waste that is spilled or dumped on the land surface. It is also the buffer zone protecting groundwater from contamination by toxic chemicals or pathogens. Thus, the vadose zone is of importance to agronomists, soil scientists, waste engineers, and a host of management professionals who are concerned with protection of groundwater. With its obvious significance to so many disciplines, we might expect that research activity in the vadose zone would have produced a thorough characterization of its important transport, retention, and reaction properties, but this is not the case. Although a number of experiments have been conducted in the vadose zone to understand water or chemical movement and fate, most studies have been monitored only in the top 1–2 m [Flury, 1996]. Moreover, few experiments have been conducted in
clay-rich soils because of the difficulties in monitoring and sampling, and the effect of soil layers on transport in any soil type is largely unknown. Even though large areas are often involved in applications of vadose zone hydrology, most of the information we have has been obtained through experiments conducted at the scale of the soil plot or smaller. Moreover, although replicated plot experiments have provided some information on the spatial variability of transport processes, little knowledge exists about the variation of these processes over time [Jury and Fluhrer, 1992].

In addition to experimental limitations, a number of theoretical gaps persist in our understanding of the vadose zone. Soil water hysteresis continues to provide dilemmas for modelers, particularly when secondary scanning loops are involved. Many coupled flow interactions (involving, for example, water and heat movement) are incompletely understood. Multiphase flow of nonaqueous phase liquids and water is an important area of research in the vadose zone, but attempts to model it have been limited by failure to represent unstable flow (among other problems). Upscaling, or the transformation of experimental or theoretical results from one space or time scale to another, is an active research area, but no consensus has been reached about the best approach to use in information transfer. Stochastic-continuum modeling, which has enjoyed success in groundwater studies, has limited application to vadose zone hydrology because its assumptions are not well met in unsaturated soil. In particular, flow in the vadose zone is normally perpendicular to the direction of stratification, so that the properties of the medium relevant to transport are often nonstationary.

Flow and transport processes in the vadose zone differ from those in groundwater in several important respects. The resistance offered by the matrix to water flow is a non-linear function of the degree of saturation or the energy state of the water. Because of the transient processes occurring at the inlet boundary, flow and transport processes in the vadose zone rarely, if ever, reach a steady state. Since temperature and air-pressure changes can significantly influence the soil environment, representation of the flow and transport regime requires multiphase characterization. And whereas in groundwater the direction of flow is generally parallel to the natural stratification, in the vadose zone water and chemicals often move perpendicular to the natural layering pattern of soil. In addition, spatial and temporal variability of important flow and transport properties is substantial, leading to extreme data demands for both deterministic and stochastic modeling. As a result, theories for flow and transport processes in the vadose zone have not evolved as rapidly as those describing movement in groundwater. At present, the Richards' flow equation is used in virtually all simulations of water flow in the vadose zone, despite its limiting assumptions, and solute transport is represented by the advection dispersion equation, which suffers from similar drawbacks.

Field experiments of flow and transport have yielded mixed support for either of these equations. Such experiments have frequently revealed the existence of varying degrees of so-called preferential flow, wherein water moves at much higher than average rates through a small portion of the soil volume. The reasons why preferential flow occurs are soil dependent, but include flow through structural voids [White, 1985], channeling caused by discrete obstacles within the soil matrix [Kung, 1990], and fluid instability, arising from a host of causes [Hillel and Baker, 1988]. Once believed to occur exclusively in structured soils, preferential flow is now recognized as prevalent under a wide range of conditions in permeable, structureless soils as well as those containing cracks and crevices [Flury, 1996]. Since the theme of this conference is flow in fractured rock, I will tailor the rest of the discussion to preferential flow in the vadose zone, which is the transport process most relevant to the discipline of fractured rock hydrology.

2. FIELD STUDIES OF PREFERENTIAL FLOW

Preferential flow has been observed by researchers and field workers for many years. However, until very recently, experimental tests designed specifically to measure properties of preferential flow have been lacking. One of the difficulties with monitoring preferential flow is finding a tracer that can be detected when preferential flow is rapid and also confined to a small portion of the soil matrix. Two strategies have dominated field research thus far: use of visible dyes that are added to the surface and later exposed by digging a trench to reveal the pattern along a vertical face, and use of a strongly adsorbed compound that will migrate under preferential flow to a depth substantially greater than it would reach by complete interaction with the entire soil matrix.

Jury et al. [1986] conducted a comprehensive field study with a strongly adsorbed pesticide (napropamide) that was added to the surface of a 1.44 ha sandy loam field. The pesticide was leached by sprinkler irrigation for two weeks until 25 cm of water had been added. Then, 19 soil core samples were taken at random locations on the field and sampled in 10 cm increments for napropamide concentration. About 72% of the recovered pesticide was found in the top 20 cm, where it would be expected to be if the water flow rate were uniform, but the remaining 28% was distributed erratically between 30 and 190 cm.
A second pesticide transport study was conducted by Ghodrati and Jury [1990] on the same field. In this study, sixty-four 1 m x 1 m plots were divided into a set of treatments that used one of four different methods of water application (continuous and intermittent sprinkling or ponding), three different methods of pesticide application (wettable powder, emulsifiable concentrate, or dissolved), and two different surface preparation methods (sifting and replacing the top 20 cm or leaving it undisturbed). Each plot received exactly 12 cm of water and pulses of three pesticides (atrazine, napropamide, and prometryn) having different mobilities, together with a mobile tracer (bromide).

At the conclusion of the water application, each plot was sampled in 10 cm increments to a 2 m depth by three soil cores that were combined prior to analysis. Preferential flow to varying degrees was observed in all plots regardless of treatment, with the sole exception that wettable powder could not penetrate the repacked soil layer. The amount of preferential flow averaged over all treatments was 22%, similar to what had been found earlier by Jury et al. [1986] on the same field. Although somewhat less preferential flow was observed for the most strongly adsorbed compound, the maximum depth of penetration into the soil was independent of the extent of adsorption potential. As part of the study, Ghodrati and Jury [1990] added a mobile red dye to the surface of the plots and excavated along a lateral trench face to observe the flow patterns. In virtually every case, the wetting front revealed leading-edge plumes that resembled instabilities, and the paths followed by the plumes appeared to lead through the same soil material as that bypassed.

Kung [1990] conducted a comprehensive dye-trace study on a field containing Plainfield sand, a highly permeable soil without pronounced layers or significant structural features. He added a pulse of the dye to the surface of a 4 m x 4 m plot, washed it in, and excavated the entire plot to a depth of 6 m. Although much of the plot area was covered with dye near the surface, the water channeled into smaller and smaller zones at greater depths, finally occupying less than 2% of the area at the bottom of the excavation. Detailed examination of the flow paths revealed that the water was flowing around discrete sand lenses imbedded in the matrix, joining with other isolated flow channels that were also meandering through the matrix. He termed the phenomenon “funnel flow.”

Other studies of pesticide transport under field conditions have reported higher than expected mobility for compounds that adsorb to the soil. In a field study on a crop silt-loam soil, Gish et al. [1986] reported that the laboratory-measured adsorption coefficient for atrazine greatly underestimated its mobility relative to bromide. Hornsby et al. [1990] monitored bromide together with aldicarb and its degradation products for 200 days under a citrus grove and sandy soil in Florida. They observed that bromide and aldicarb had comparable mobility, even though the latter compound was predicted to have a retardation factor between 1.1 and 2.0, based on laboratory measurements. Because of the large amount of rainfall in this area, both bromide and aldicarb reached groundwater at 7.2 m during the 200 day monitoring period.

Bowman and Rice [1986] conducted a large study on a sandy loam field in which a water tracer and bromacil, a mildly adsorbed herbicide, were applied and leached under periodic ponding. Nearly all the individual soil cores showed the bromacil to be retarded with respect to the water tracer, but both were moving faster than predicted by piston flow. Tile drains have been used to monitor nutrients and pesticide-leaching losses below cropland for many years. Recently, they have been used to detect the extremely high mobility of dissolved chemicals. Richard and Steenhuis [1988] reported chloride arrival at 80 cm in the tile-drain effluent of a sandy loam, with the first outflow following application of the chemical to the surface. Subsequent rainfall events also triggered additional outflow pulses of chloride. A similar phenomenon was reported by Kladivko et al. [1991] on a tile-drain silt-loam soil. Following a single application to the surface, traces of the pesticides appeared in the tile effluent at 0.75 m after only 2 cm of net drainage. These pulses tapered off before the water quit flowing out of the tile, but returned with subsequent rainfall events through the season.

In the majority of these studies, preferential flow was a significant but not major part of the total flow regime. In more heterogeneous soil, however, it has been observed to dominate. Roth et al. [1991] performed a field study on a layered soil with a highly variable texture and structure and found that preferential flow was responsible for about 55% of the total mass transport through the top 2 m. Flury et al. [1994] conducted comprehensive dye-trace studies on the 14 major agricultural soil types in Switzerland and concluded that preferential flow was a significant or dominant feature of the flow regime in all but one of them. In a subsequent review article that surveyed all existing pesticide transport studies in unsaturated field soil, Flury [1996] concluded that preferential flow of even strongly adsorbed compounds was commonly observed in structureless soils, particularly loamy textures, without any apparent cause.

3. CAUSES OF PREFERENTIAL FLOW

Despite all of the evidence and effort at monitoring, no correlation has been observed between the local values of
soil properties and the location where preferential flow occurs in structureless, vertically homogeneous soil. Various mechanisms have been postulated, including water repellency [van Dam et al., 1990], air entrapment [Peck, 1965], and small-scale (i.e., 1 cm) variations in soil properties [Roth, 1995]. But definitive proof relating the pathways, intensity, and extent of preferential flow to measurable soil matrix properties has not been found. One reason why this might be true has been largely overlooked: the preferential flow may be the result of an instability in the flow field that is not caused exclusively by local permeability variations. Furthermore, the instability may be a consequence of one of the most common hydrologic events in soil: redistribution following the cessation of rainfall or irrigation. Early theoretical analyses by Raats [1973] and Philip [1975], using the Green and Ampt infiltration model, showed that redistribution was inherently unstable, even in homogeneous soil, and that any deviation from perfect one-dimensional advance of the wetting front would grow into a finger. The condition responsible for this instability is the matric potential gradient behind the wetting or draining front: whenever it opposes the flow, instabilities will develop out of perturbations in the uniformity of the wetting front as it advances into the soil.

This conclusion was tempered somewhat by Diment et al. [1982]. They conducted a simplified stability analysis of the Richards' equation in homogeneous soil and showed that instabilities would only develop if the initial water content of the soil ahead of the advancing front were extremely dry. They also demonstrated that the tendency toward instability increased as the wavelength of the disturbance increased, implying that the unrestricted width of the field environment may be an important factor in the development of instability. Their analysis was limited to a restricted set of soil conditions (perturbed one-dimensional infiltration and redistribution into homogeneous soil with uniform initial wetting) and contained some simplifications (no hysteresis, uniform velocity of wetting front).

Laboratory validation of unstable flow in homogeneous soil has been demonstrated repeatedly in Hele-Shaw cells, in which redistribution follows infiltration into oven- or air-dry soil [Glass et al., 1989; Selker et al., 1992; Wang, 1997]. However, Diment and Watson [1985] demonstrated that redistribution became stable in their Hele-Shaw cell when the initial water content of the soil was increased by only a few percent. The reason for this is that the wavelength of the fingers becomes large enough that the side walls of the column will not permit an instability to propagate. This observation also offered an explanation as to why preferential flow is so rarely observed in the laboratory. If the dimension of the column perpendicular to the direction of flow is not large compared to the wavelength of the unstable front, a finger will not propagate because it will not have sufficient lateral flow occurring to feed the advance of the instability.

Unstable flow is considerably more likely to occur when soil conditions deviate from ideality. Increasing air compression ahead of an advancing wetting front will induce instability at the time when the soil water potential (matric + air pressure) at the front exceeds that above it [Peck, 1965; Parlange and Hill, 1976]. Abrupt soil layering, with a finer-textured soil above a coarser textured one, will interrupt flow until the pressure reaches a critical point, at which time infiltration will advance into the lower zone at one particular location [Hill and Parlange, 1976]. Furthermore, discrete soil lenses of either coarser or finer soil than the matrix surrounding them can cause channeling. This channeling will induce a preferential flow in part of the soil volume as long as the matrix conductivity is large enough to support it [Kung, 1990]. If preferential flow in structureless soils is caused by fluid instability rather than simply being a consequence of variations in local soil properties, the implications for water flow and chemical transport research and management are profound. Simply by virtue of their widespread use in research and management, water flow models have gained an acceptance for simulation of flow under field conditions that arguably is not warranted by their record of achievement. They either ignore the possibility of preferential flow entirely by using the volume-averaged Richards' equation to model water movement, or introduce parameter-intensive methods of inducing preferential flow entirely as a result of soil property variations.

If preferential flow is initiated and reinforced by fluid instability rather than property variability, models that use the Richards' equation will not predict its occurrence with their current approach. Despite the wide variation in soil conditions and experimental design, field studies of preferential flow have all shared one common feature: the time of sampling was always scheduled 24 hr or more after cessation of water application, allowing substantial time for redistribution. According to theory, redistribution can favor the propagation of an instability should any perturbation occur in a sharp draining front. Since numerous local features in a heterogeneous field soil could create a perturbation in an advancing wetting front, efforts to characterize the exact location where the process starts, in terms of a measurable soil property, are unfeasible with current technology. In the
interim, a more practical and fruitful approach might be to develop an understanding of the water application conditions that will turn a flow perturbation into an instability.

4. CONCLUDING REMARKS

Improved understanding of field-scale preferential flow will emerge from new developments in monitoring and more experimental testing under natural conditions. At the moment, we do not have answers to a number of basic questions that limit our ability to manage or model this phenomenon: Which measurable soil properties are important in predicting the onset of preferential flow? Which measurable soil properties are important in predicting the extent of preferential flow? How fast do water and dissolved chemicals move in a preferential flow channel? How can preferential flow be enhanced or reduced in a given environment? What is the role of a water application method or surface water management in producing preferential flow? To what extent do sorption reactions occur during preferential flow, and how can they be characterized as a function of measurable soil properties?

The standard tools for monitoring water flow in soil are inadequate for addressing these questions. Field monitoring of preferential flow has thus far been possible only using dyes, densely replicated soil cores, or analysis of tile-drain effluent. These tools either fail to characterize the speed of preferential flow, as is the case with dyes or coring, or provide no spatial resolution of the preferential flow event, as with tile-drain monitoring. To make real progress in characterizing preferential flow, we need rapid-response tensiometers and solution samplers that offer minimum disturbance of the soil, and we need some means of monitoring a substantial portion of the soil volume at a high degree of spatial resolution.

REFERENCES


William A. Jury and Zhi Wang, University of California, Riverside, CA 92521.